2022

Level 2 General

Biostatistics

Mid-term

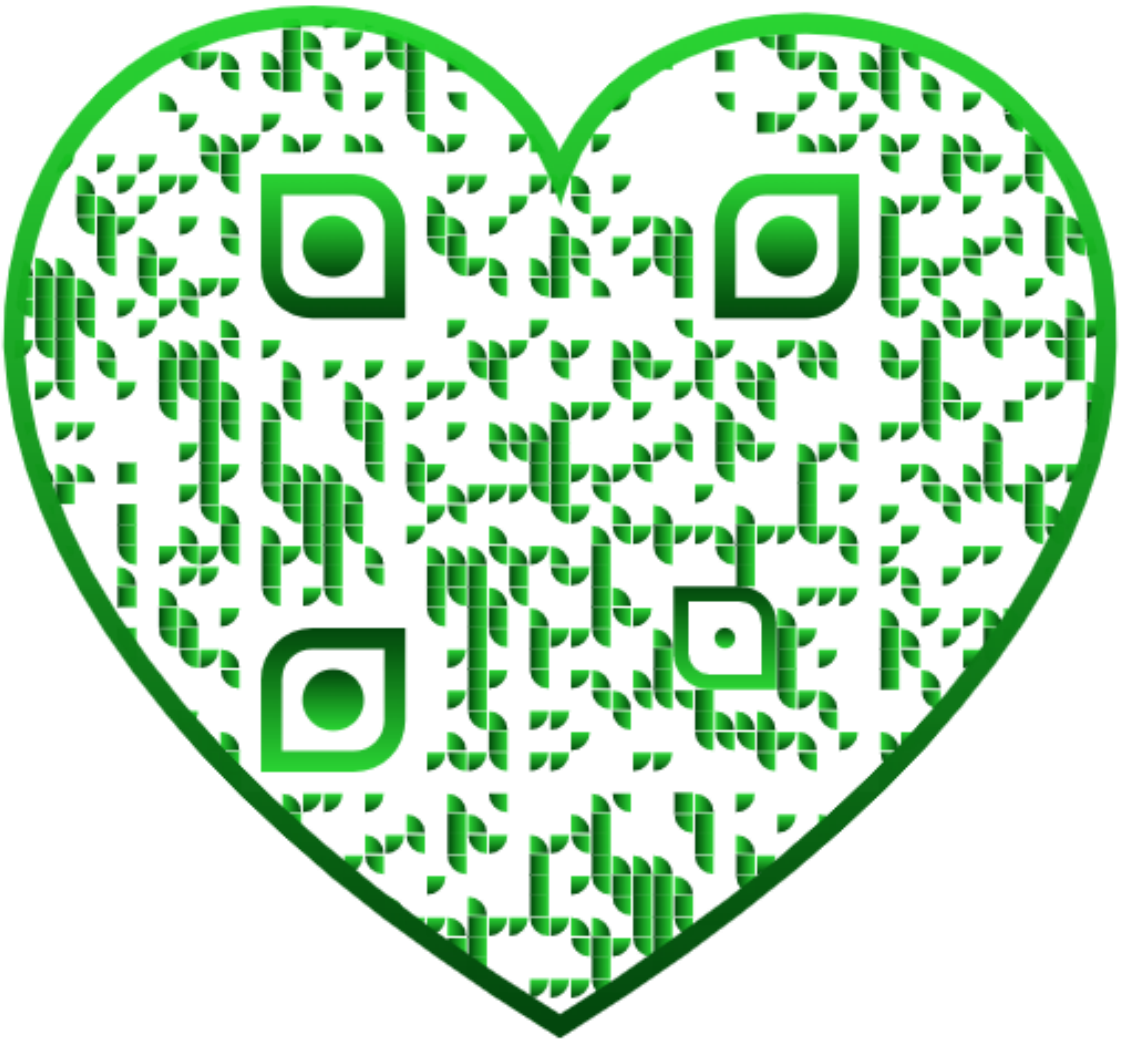
Complete part

Dr Hoda Ezzat

Prepared By

Dr.Ahmed Elshazly

01068865764



Biostatistics

Statistics	Biostatistics	Data
❑ **Science of collecting, summarizing**, presenting and analyzing data ⇨ This leads to conclusion and subsequent decision.	❑ When the data analyzed are derived from the **biological science and medicine**, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.	❑ The raw material of statistics is data. ❑ We my define data as figures ⇨ results from the process of: ↳ Counting (Hospital administrator counts the number of patients) or ↳ Measurement (A nurse weighs a patient).

Source of Data

We search for suitable data as the raw material for our investigation ⇨ Such data available from one or more of the following sources.

1] Records (routinely kept)	2] External sources	3] Survey	4] Experiments
❑ Hospital medical records contain immense amounts of information on patients. ❑ Hospital accounting records contain a wealth of data on the facility's business.	❑ The data needed to answer a question may already exist in the form of published reports, commercially available data banks, or the research literature. ❑ Ex: someone else has already asked the same questions.	❑ The source may be survey, if the data needed is about answering certain questions. ❑ Ex: If the administrator of the clinic wishes to obtain information regarding the mode of transportation used by patients to visit the clinic, then a survey may be conducted among patients to obtain this information.	❑ Frequently the data needed to answer a question are available only as the result of an experiment. ❑ Ex: If a nurse wishes to know which of several strategies is best for maximizing patient compliance, she might conduct an experiment in which the different strategies of motivating compliance are tried with different patients.

Some important terms

A Variable	Population	Sample
❑ It is a characteristic that takes on different value in different persons, places or things such as: ❶ Heart rate ❷ height of adults males. ❸ The weights of preschool children. ❹ The ages of patients seen in a dental clinic.	❑ The largest collection of value of a random variable for which we have an interest at a particular time. ❑ Ex: weight of all the children enrolled in a certain elementary school. ❑ Population may be finite or infinite.	❑ It is a part of population. ❑ The weight of only a fraction of these children.

Types of variable

Quantitative variables

- ❑ It can be measured in usual sense.
- ❑ Ex: the heights of adult males – the weight of preschool children – ages if patients seen in dental clinic.
- ❑ Types:

Discrete	Continuous
❑ Characterized by gaps or interruptions in the value that it can assume ⇨ such as: 1) No of admission to a general hospital. 2) No of decayed, missing or filled teeth per child.	❑ Can assume any value within a specified relevant of values assumed by the variable ⇨ such as: Height – weights – skull circumference. ❑ No matter how close together the observed heights of two people, we can find another person whose height falls somewhere in between.

Qualitative (Categorical) variables:

- ❑ Many characteristics are not capable of being measured.
- ❑ Types:

Ordinal	Nominal
Can be recorded such as Classification of people into socio-economic groups.	Cannot be recorded such as hair color.

Role of Biostatistics

1) To guide the design of an experiment or survey prior to data collection.
2) To analyze data using proper statistical procedures and techniques.
3) To present and interpret the results to researchers and other decision makers.

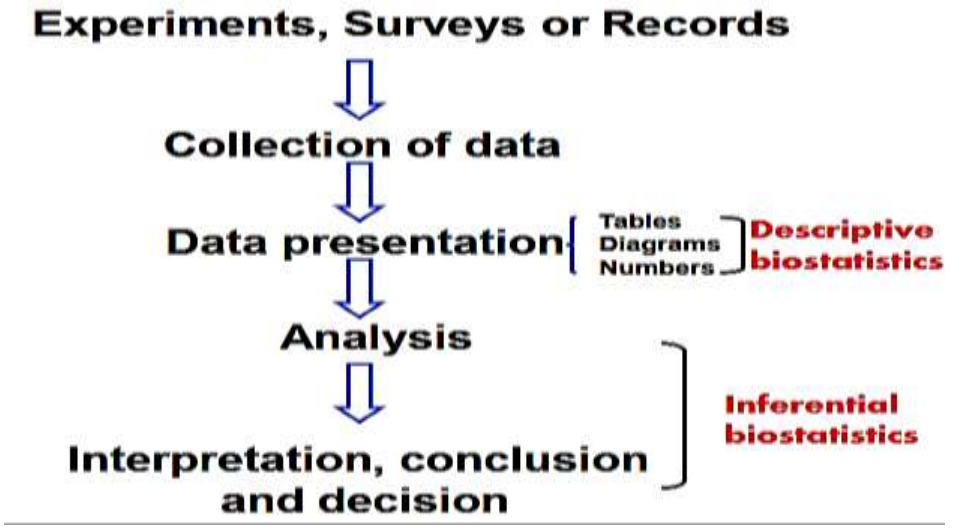
Classification of Biostatistics

1] Descriptive Biostatistics

- ❑ Mainly description of data.
- ❑ No attempt to draw conclusion from this information.
- ❑ They are divided into:
  - a) Tabular presentation of data.
  - b) Diagrammatic presentation of data.
  - c) Numerical presentation of data.

2] Inferential Biostatistics

- ❑ A decision, estimate, prediction or generalization about a population, based on a sample.



1] Descriptive Biostatistics

Descriptive هنتكلم فى الصفحات الجايه عن ال

A] Tabular presentation of data

1] Frequency distribution table

For Discrete random variables

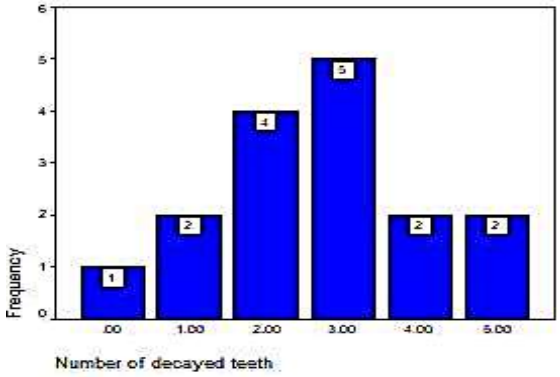
- Example: suppose that we take a sample of size 16 from children in 1ry school and get the following data about decayed teeth.

  3, 5, 2, 4, 0, 1, 3, 5, 2,3, 2, 3, 3, 2, 4, 1.

- To construct a frequency table:
  1) Order the values from smallest to largest.

     0, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 5, 5
  2) Count how many numbers are the same.

No of decayed teeth	Frequency	Relative Frequency
0	1	0.0625
1	2	0.125
2	4	0.25
3	5	0.3125
4	2	0.125
5	2	0.125
Total	16	1



Representing the simple frequency dist. Table using the bar chart

For Continuous random variables

- For large samples, we cannot use the simple frequency to represent the data.
- We need to divide the data into groups or intervals or classes.
- So, we need to determine:

1) The number of intervals (K)

- Too few intervals are not good because information will be lost.
- Too Many intervals are not helpful to summarize the data.
- A commonly followed rule is that: $6 \leq K \leq 15$ or $K = 1 + 3.322\,(\log n)$.

2) The range (R)

- It is the difference between the largest and smallest observation in the data set.

3) The width of the interval (w)

- Class intervals generally should be one of the same width, thus if we want K intervals, then W is chosen such that: $W \geq R/K$

Example 1

- We wish to know how many class intervals to have in the frequency distribution of data in table.
- 9 – 10 of ages of 189 subjects who participated in a study on smoking cessation.

Solution:

Since the number of observation is 189, then:

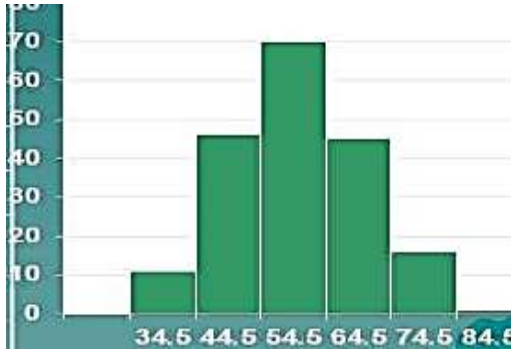
- K = 1 + 3.322 (log189) ≅ 9
- R = 82 – 30 = 52
- W = 52 / 9 = 5.5778
- It is better to let W = 10, then the intervals will be in the form:

Class interval	Frequency
30 - 39	11
40 - 49	46
50 - 59	70
60 - 69	45
70 - 79	16
80 - 89	1
Total	189 (sample size)

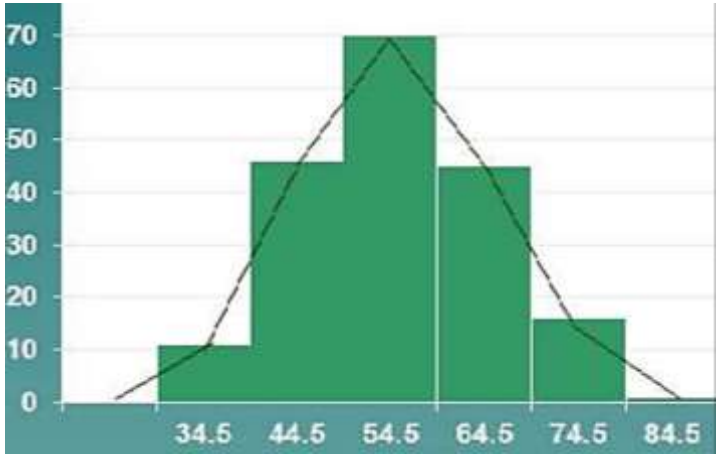
Rrpresenting the grouped frequently table using histogram

- To draw the histogram, the true classes limits should be used.
- They can be computed by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit for each interval.

Class interval	Frequency
29.5 - < 39.5	11
39.5 – < 49.5	46
49.5 – < 59.5	70
59.5 - < 69.5	45
69.5 - < 79.5	16
79.5 – < 89.5	1
Total	189 (sample size)



Rrpresenting the gp frequently table using histogram



Example 2

- The systolic blood pressure in mm/Hg of 50 healthy individuals.

122 – 112 – 132 – 122 – 120 – 126 – 118 – 110 – 120 – 125 – 119 – 120 – 116 – 120 – 124 – 120 – 123 – 135 – 121 – 132 – 120 – 115 – 122 - 112 – 128 – 122 – 118 – 116 – 129 – 121 – 121 – 122 – 114 – 130 – 120 – 116 – 134 – 118 – 126 – 120 – 116 – 134 – 128 – 126 – 122 – 126 – 122 – 120 – 126 – 118

1) Arrange data values in an ascending way:

110	116	118	120	120	122	122	125	126	132
112	116	118	120	120	122	122	126	128	132
114	116	119	120	121	122	123	126	128	134
115	118	120	120	121	122	124	126	129	134
116	118	120	120	121	122	125	126	130	135

2) Calculate the range = the highest value - the lowest value = 135 – 110 = 25.

3) Decide the number of classes desired.

4) Determine the class interval = **range/number of classes** = 25/5 = 5

5) Midpoint of the class = **Lowest value in the class + 1/2 class interval**

6) Frequency distribution table:

Class number	Class-	Midpoint of the class	Frequency
1	110-	112.5	3
2	115-	117.5	10
3	120-	122.5	22
4	125-	127.5	9
5	130-	132.5	6
			Total = 50

- ***Conclusion***: The most frequent BP of individuals is that lying in the 3rd class (120-125 mm/Hg) with a midpoint 122.5 mm/Hg.

Example 3

The marks of 52 students in a biostatistics exam were as follow:

2 –2 –5 – 1 – 2 – 6 – 3 – 3 – 4 –2 – 4 – 0 – 5 – 7 – 7 – 5 – 6 –6 –8 – 10 – 7 –2 – 2 - 10 – 5 – 8 – 2 –5 – 4 –2 – 6 –2 – 6 – 1 – 7 – 2 – 7 – 2 – 3 – 8 – 1 – 5 – 2 – 5 –2 – 14 – 2 – 2 – 6 – 3 – 1 - 7

1) Arrange data values in an ascending way:

0	2	2	2	3	4	5	6	7	7	10
1	2	2	2	3	4	5	6	7	8	14
1	2	2	2	3	5	5	6	7	8	
1	2	2	2	3	5	5	6	7	8	
1	2	2	2	4	5	6	6	7	10	

2) Calculate the range = the highest value – the lowest value = 14 – 0 = 14

3) Divide on the number of the classes (categories) desired (5 classes).

4) Determine the class interval = range/number of classes = 14/5 = 2.8 ≈ 3

Frequency distribution table:

Class	Frequency
0-	20
3-	14
6-	15
9-	2
12-	1
	Total = 52

2] Cumulative Frequency Distribution table "more descriptive"

- ❑ Used when it is required to know how many individuals have a value less or more than a certain value.
- ❑ The values are arranged in ascending manner (ascending cumulative frequency tables) or descending manner (descending cumulative frequency tables).

Ascending cumulative frequency distribution table

Start with the second class:

Systolic blood pressure (mm/Hg)	Number of individuals	Percent of the total
Less than 115	3	6
Less than 120	13	26
Less than 125	35	70
Less than 130	44	88
Less than 135	50	100

Descending cumulative frequency distribution table

Start with the class just before the last one.

Systolic blood pressure (mm/Hg)	Number of individuals	Percent of the total
More than 130	6	12
More than 125	15	30
More than 120	37	74
More than 115	47	94
More than 110	50	100

B] Graphical presentation

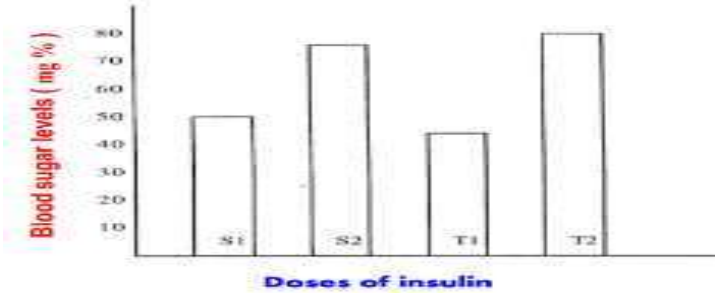
1] Bar chart

- It is a graphical presentation of the data by rectangular of constant width, drawn with length proportional to frequency, relative frequency or magnitude.
- Bar charts used to depict categorical information e.g., Male–Female, Mild–Moderate–Severe, etc.

A] Simple bar chart

- If we are representing the data of one phenomenon a simple bar chart is used as in the following:
- The blood sugar levels of rabbits after treatment with 2 doses of the standard and 2 doses of the test solution of insulin.

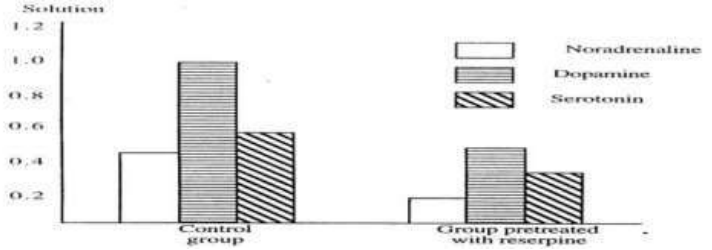
Dose of insulin	S1	S2	T1	T2
Blood sugar mg%	50	75	42	80



b] Multiple bar chart

- A chart depicting 2 or more characteristics in the form of bars of length proportional in magnitude of the characteristics.
- The effect of pretreatment with reserpine on concentration of noradrenaline, dopamine and serotonin in the brain of rats.

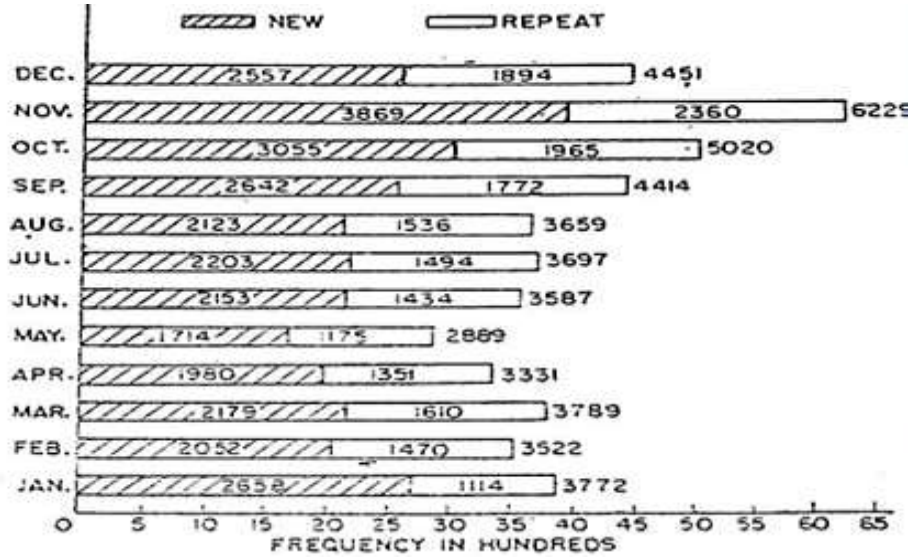
Amine	Control group	Pretreated group with reserpine
Noradrenaline	0.4	0.16
Dopamine	1.06	0.50
serotonin	0.54	0.30



C] Component bar chart

- Example: Monthly distribution of new and repeat patients in a hospital during the year 1973.

Month	Number of cases		
	New	repeat	Total
January	2658	1114	3772
February	2052	1470	3522
March	2179	1610	3789
April	1980	1351	3331
May	1714	1175	2889
June	2153	1434	3587
July	2203	1494	3697
August	2123	1536	3659
September	2642	1772	4414
October	3055	1965	5020
November	3869	2360	6229
December	2557	1894	4451



2] Pie (Sector diagram)

- Another way of presenting data of **qualitative characters** such as blood groups, Rh groups, age groups, sex groups etc.
- The frequencies of the group are shown in a circle [Degree of angle and area of the sector ➔ indicate the frequency].
- Ex: The following is the distribution of blood groups in certain city:

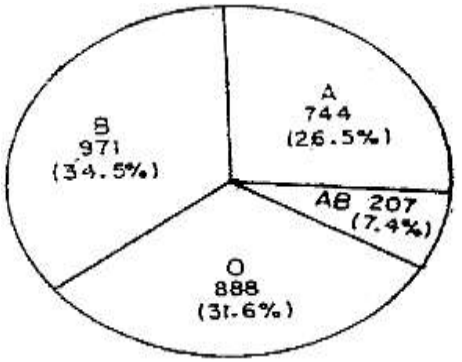
Blood group	Frequency
A	744
B	971
O	888
AB	207
Total	2810

Blood group	Number	Percent	Degree
A	744	26.5	95.4
B	971	34.5	124.2
O	888	31.6	113.8
AB	207	7.4	26.6
Total	2810	100	360

- Size of each angle is calculated by multiplying the class percent by 360
  Angle = (% of class/100) X 360
- Or

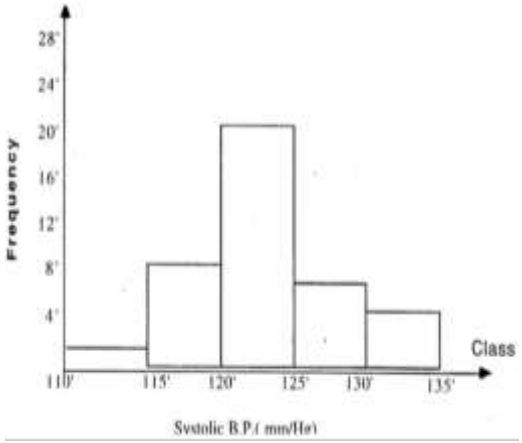
$$\frac{Class\ frequency}{Total\ observation}\ x\ 360^{o}$$

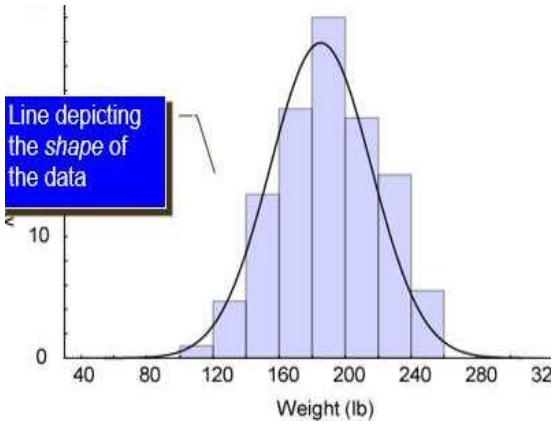
- Size of angle for blood group A =
  = 26.5 x 360/100 = 95.4
  Or 744/2810 X 360 = 95.4



3] Histogram

- It is a **continuous bar chart** presenting data in a frequency distribution.
- **Distributions are often symmetrical** with most scores falling in the middle and fewer toward the extremes.
- Most biological data are symmetrically distributed and form a normal curve (a.k.a, bell-shaped curve).
- The class intervals are in the X-axis, on each class a rectangle is drawn with a corresponding height to the frequency (F) on the Y-axis.

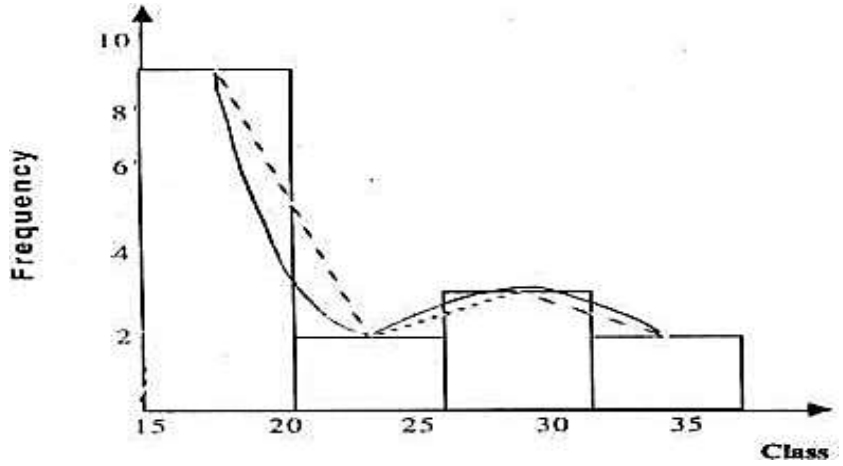




4] Frequency polygon

- If in histogram a dot is put in the middle of the upper base of each rectangle and these dots are connected by a straight line, the frequency polygon is obtained.
- Each point in this line represents the midpoint of the class and its frequency.

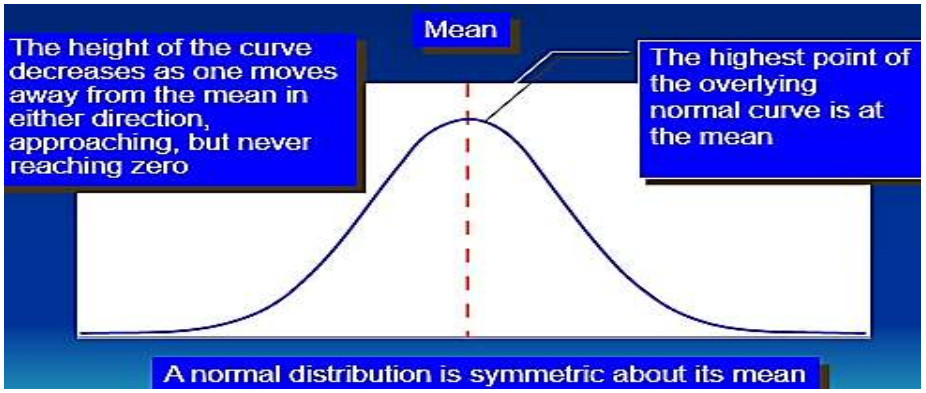
Frequency polygon (the dotted lines)



5] Frequency distribution curve

- If frequency polygon is smoothed, the frequency distribution curve is obtained.
- Or by direct plotting of the frequency against the distribution of a certain phenomenon.

A] symmetric	B] Asymmetric
❑ When curve can be divided into 2 equal parts by a line passing through the centre of distribution. 1) Bell shaped (Majority). 2) U-shaped (data concentrate at both ends).	❑ When the distributions extend further in one direction than the other. 1) Positively skewed (extremes in the higher values shifting the curve towards the right). 2) Negatively skewed (less common, extremes in the lower values shifting the curve towards the left).


Mean
The height of the curve decreases as one moves away from the mean in either direction, approaching, but never reaching zero
The highest point of the overlying normal curve is at the mean
A normal distribution is symmetric about its mean

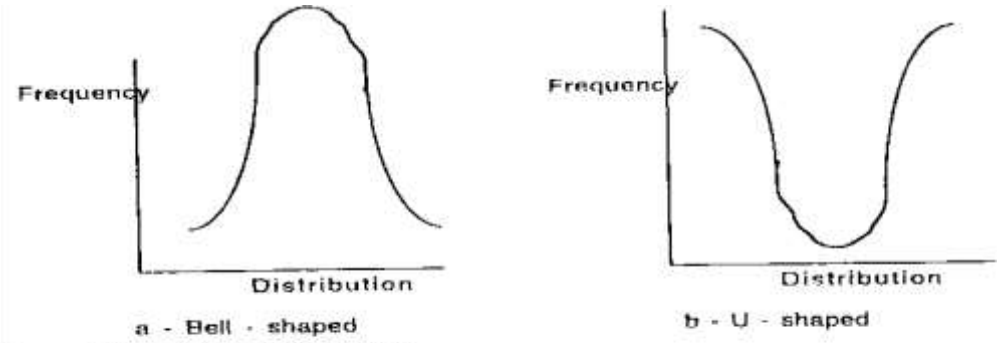
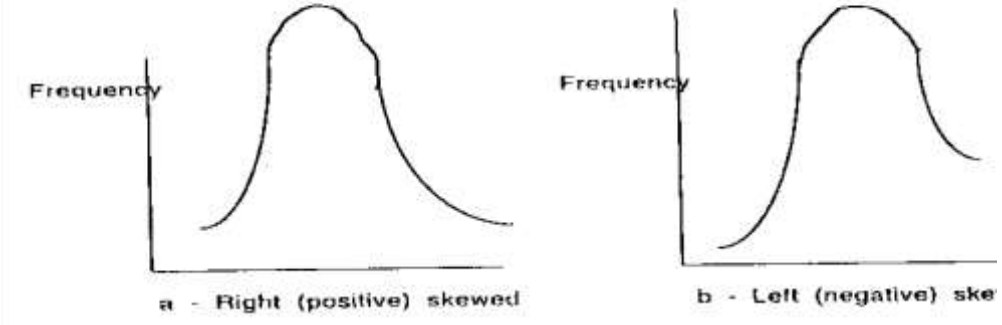
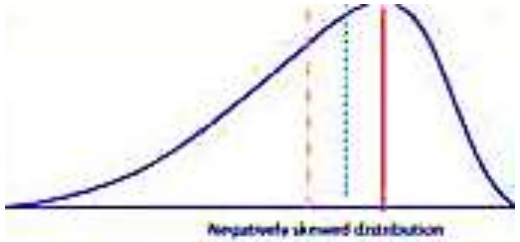


Figure (3) Symmetic distribution curves .



- The data are not didtributed symmetrically in skewed distrubutions.
- Consequently, the mean, median and mode are not equal and are in diiferent positions.
- Scored are clustered at one end of the distrubitions.
- A small number of extreme values located in the limits of the oopsite end.
- ecause the mean is shifted so much, it is not the best estimate of the qverage score for skewed distrubtions.
- The median is a better estimate of the center of skewed distrubtions.
  - ↳ It will be the central point of any distubutions.
  - ↳ 50% of the values are above and 50% below the median.



Positive if skewed to right.

Negative if skewed to left.

6] Pictogram (Picture diagram)

- It is a diagrammatic representation of frequencies, R.F. or magnitude using symbols e.g. drawing a picture.
- **Example:** The pictures are drawn in horizontal lines.
- Each picture indicates a unit of 10, 20, 100, 1000 or 10,000 happenings. Number of pictures in each row gives a quick idea of relative frequency of an attribute. Fraction of a unit has to be ignored, though half (0.5) may be indicated by half picture as in Fig. (3), which gives incidence of cholera cases in 4 cities.

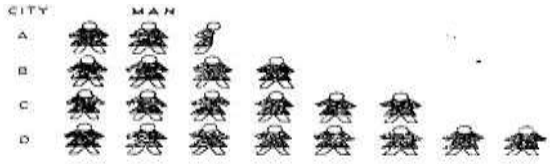


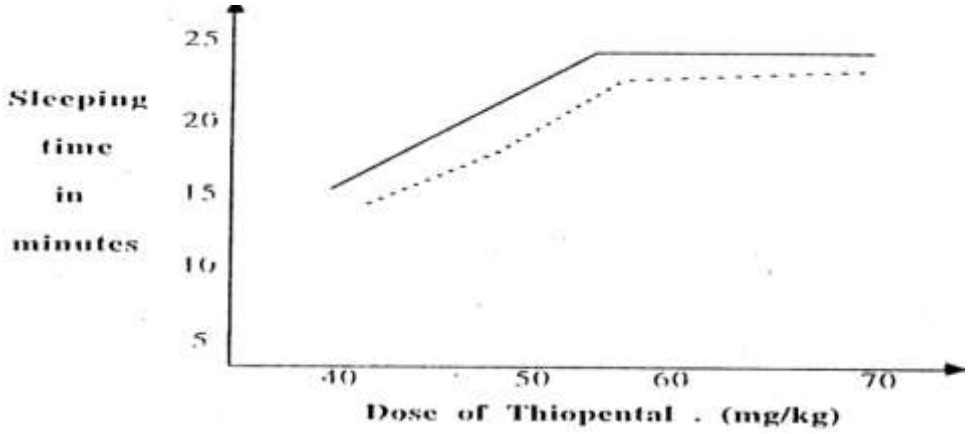
Figure 12: Pictogram of cholera in the 4 cities, each man's figure indicates 10 cases.

9] Line graph

A] Natural or Arithmatical

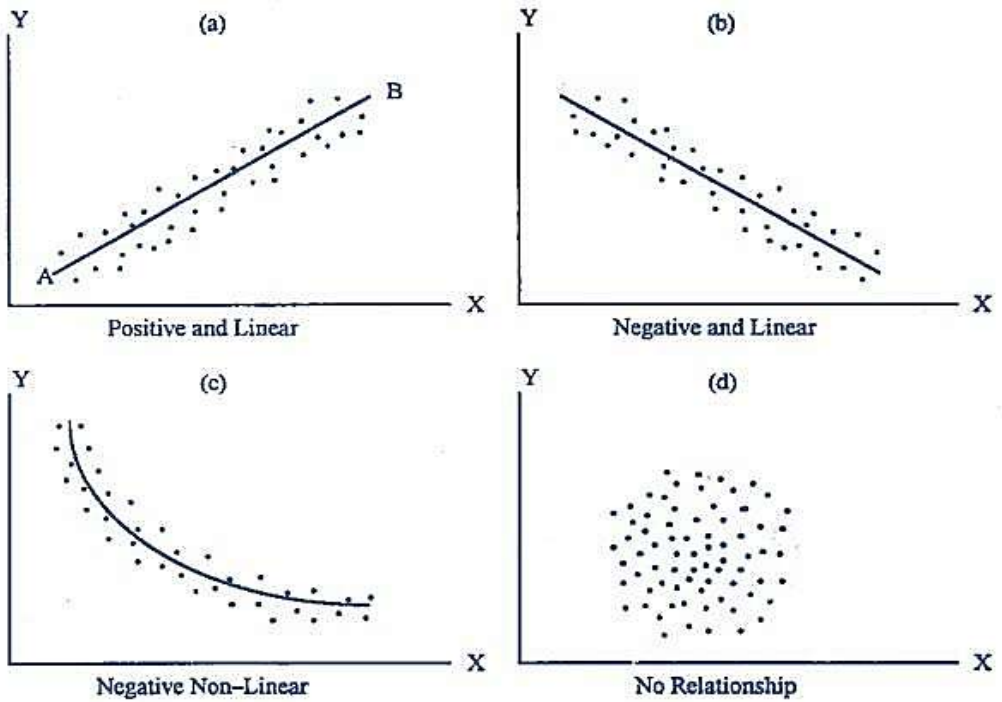
- In presentation of data by the use of graphs, lines are drawn perpendicular to one another, and are divided to scales of values.
  - The horizontal line is known as the X-axis or the abscissa.
  - The vertical line is known as the Y-axis of the ordinate.
- If we are presenting more than one phenomenon but on one and the same variable ➔ more than one line should be drawn **(lines should be differentiated by coloring or dashing or any other way).**
- Example:
  - The sleeping time of mice was determined after an I.V. Injection of different doses of thiopental to 2 groups of mice.
  - The 1st group was pretreated with chlorpromazine while the 2nd was control.

Dose of thiopental (mg/Kg)	Pretreated mice	Control mice
40	15	10
50	18	12
60	22	20
70	25	22



B] Scatter or Dot diagram

- It shows the nature of correlation between two variable characters X & Y in the same persons or groups such as height and weight in men aged 20 years ➔ So, called [Correlation diagram].
- The characters are read on the base (height) and vertical (weight) axes and the perpendicular drawn from these readings meet to give one scatter point.
- Varying frequencies of the characters give a number of such points or dots that show a scatter.
- A line is drawn to show the nature of correlation between variables.



Applications of Biostatistics

➲ Medical statistics: Go under different names when applied in different fields such as:

Health statistics	Vital statistics	In medicine
❑ In public health or community health.	❑ In demography pertaining to vital events of birth, marriage and death.	❑ It related to the study of defect, injury, disease, efficacy of drugs, serum level and line of treatment etc...

Application and uses of Biostatistics as a science

1] In physiology and anatomy	2] In pharmacology	3] In medicine
A) To define what is normal or healthy in a population and to define limits of normality in variables as weight and pulse rate. ↳ The mean pulse rate is 72/min but up to limits it may be normal on either side of mean has to established with certain appropriate techniques. B) To find the difference between means and proportions of normal at 2 places or in different periods. C) To find the correlation between two variables X and Y such as height and weight.	A) To find the action of a drug. A drug is given to animals or human to sees whether the changes produced are due to the drug or by chance. B) To compare the action of two different drugs or two successive dosage of the same drug. C) To find the relative potency of a new drug with respect to a standard drug.	A) To compare the efficacy of a particular drug, operation or line of treatment. For this the percent cured, relieved or died in experiment and control groups is compared and find if the difference is due to chance or not. B) To find an association between two factors such as smoking and lung cancer. C) To identify signs and symptoms of a disease or syndrome e.g. cough in typhoid.

Notations for a population and sample values

A Statistic	A Parameter
➲ It is a descriptive measure computed from the data of a sample.	➲ It is a descriptive measure computed from the data of a population. ➲ Since it is difficult to measure a parameter from the population, a sample is drawn of size n, whose values are $\chi 1$, $\chi 2$, ..., χn. ➲ From this data, we measure the statistic.

Roman letters are used for statistics of sample & Greek for parameters of population, common notations are

Value	Sample	Population
Mean	$\bar{X}$	μ
Standard deviation	S	δ
Variance	S^2	δ^2

Measures of central tendency (Averages)

➲ Is a measure which indicates the location of the middle of the data.

➲ The three most commonly used measures of central tendency are: 1] Arithmetic Mean ($\bar{X}$). 2] Median (Middle). 3] Mode.

1] Mean or Arithmetic Mean

Law

Mean	Advantages (Properties)	Disadvantages
$X^- = \frac{\sum x}{N}$	➲ Uniqueness ⇨ For a given set of data there is one and only one mean. ➲ Simplicity ⇨ It is easy to understand and to compute.	➲ Affected by extreme values. ➲ Since all values enter into the computation.

X^- = mean. X = individual result of observation. N = number of observation.

Problems

❑ The blood glucose level of 6 normal rabbits was found to be 80-90-95-85-73-87

❑ Calculate the mean blood glucose??

X^- = (80 + 90 + 95 + 85 + 73 + 87)/6 = 85

❑ Calculate the mean of the following data: 41-36-22-29-47

X^- = (41 + 36 + 22 + 29 + 47) / 5 = 35

Note 1

❑ The **algebraic sum for deviations** of measurements of a series of observation from their mean = zero

X	X^-	d
41	35	+6
36	35	+1
22	35	-13
29	35	-6
47	35	+12
		Zero

Note 2 لم تشرح

❑ *The arithmetic mean can be calculated from:*

An assumed mean using deviation.

$$X^- = Xa \frac{\sum d}{N}$$

- Xa =assumed mean.
- d = deviation
- N = number of observation

X	Xa	d
41	30	+11
36	30	+6
47	30	+17
22	30	-8
29	30	-1
		+34 -9 = +25

X^- = 30 + 5 = 35

Calculation of X^- from frequency distribution table

Assume that in each class ➔ all sample values = midpoint of the class

Systolic B.P.	Midpoint (X)	Frequency (F)	FX
110-	112.5	3	337.5
115-	117.5	10	1175.0
120-	122.5	22	2695.0
125-	127.5	9	1147.5
130-	132.5	6	795.0
			6150

$\sum F.X = 6160$

N = 50

$X^- = \frac{\sum FX}{N}$

$\frac{6150}{50} = 123$

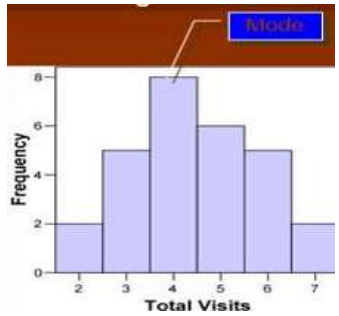
2] Median (middle)

- When ordering the data, it is the observation that divide the set of observation into two equal parts such that half of the data are before it and the other are after it

If n is odd number	If n is even number
Median is the middle of the observation. $Median\ position = \frac{N+1}{2}$	There are 2 middle observation, the median will be the mean of the 2 middle observation. $Median\ position = \frac{N}{2}, \frac{N}{2}+1$
Ex: when n = 11, the median is the 6^{th} observation.	Ex: when n = 12, the median is the 6.5^{th} observation which is an observation halfway between 6^{th} and 7^{th}

3] Mode

- It is the data value which occurs most frequently.
- The modal value is highest bar in a histogram.
- It is not always unique.
- A data set may be: Bimodal, Multimodal or no mode.
- The only measure of central tendency that can be used with nominal data.
- It is used desorbing qualitative data.



Properties:

1) **Uniqueness:** for a given of data there is one and only one mean.
2) **Simplicity:** easy to calculate.

- Ex: The heights of 5 girls are as follow:

144–102–123–162–104

Solution:

Arrangement: 102 – 104 – 123 – 144 – 162

Median position = $\frac{5+1}{2}$ The third value = 123

- Determine the mode of the following set of data:

5.4 – 1.1 – 0.42 – 0.73 – 0.48 – 1.1

Mode = 1.1

- Determine the mode of the following set of data:

A] 27 – 27 – 27 – 55 – 55–55–88–88-99

Mode = 27 and 55 (Bimodal)

B] 1 – 2- 3 – 6 – 7 – 8 – 9 – 10 [No mode]

Mode of F.D.T

The mode is the midpoint of the class having the highest F.

Heights (cm)	F
120-	10
125-	30
130-	40
135-	15
140-	5

Heights (cm)	Mid-point	F
120-	122.5	10
125-	127.5	30
130-	132.5	40
135-	137.5	15
140-	142.5	5

The mode is 132.5

Problem

Problem 1	Problem 2
1] Liver microsomal protein concentration of 8 animals were: 23–16–18–20–16–19–21–22 mg/g liver weight.	2] Find the mean, median and mode for the following list of values: 13-18-13-14-13-16-14-21-13

Measures of central tendency (Averages)

- Each of the three methods of measuring central tendency has a certain advantages and disadvantages.
- Which method should be used?
- It depends on the type of data that is being analyzed eg: categorical, continuous, and level of measurement that is involved.

Levels of measurement

- NOIR is Mnemonic to help remember the names and order of the levels of measurement.

N ⇨ Nominal O ⇨ Ordinal I ⇨ Interval R ⇨ Ratio

Measurement scale	Permissible mathematics operations	Best measure of central tendency
Nominal	Counting.	Mode.
Ordinal	Greater or less than operations.	Median.
Interval	Addition and subtraction.	Symmetrical – mean – skewed – median.
Ratio	Addition, subtraction, multiplication and division.	Symmetrical – mean – skewed – median.

Measures of dispersion

- A Measure of dispersion conveys information regarding the amount of variability present in a set of data.
- Variability is the extent values are spread above and below middle values.

Note

- If all the values are the same ⇨ there is no dispersion.
- If all the values are the different ⇨ there is a dispersion.
- If the values close to each other ⇨ the amount of dispersion is small.
- If the values are widely scattered ⇨ the dispersion is greater.

Measures of dispersions are:

1] Range 2] Variance
3] standard deviation. 4] coefficient of variation (CV)

2] Measures of dispersion

The mean is not an accurate indication (not give all information about data), We have two sets of data having the same mean ($\bar{X}$), but there is a great deviation between them.

X	$\bar{X}$	d
18	19	1
17	19	2
22	19	3
19	19	-
$\bar{X} = \frac{\sum x}{4} = 19$		
$\sum d = 1 + 2 + 3 = 6$		

X	$\bar{X}$	d
26	19	7
11	19	8
17	19	2
22	19	3
$\bar{X} = \frac{\sum x}{4} = 19$		
$\sum d = 7 + 8 + 2 + 3 = 20$		

- So the deviation of the second set of data is much greater than in the first one.
- As the deviation ↑ ➔ the mean is not representative of the data (invalid) and vice versa.

1] Range

$$R = X_{max} - X_{min}$$

➲ Disadvantages: it depends only on 2 values.

Determine the range of the following sets of data.

1] 18-17-22-19 — R1 = 22-17 =5

2] 26-11-17-22 — R2 = 26-11 =15

2] Average Deviation (A.D.)

$$A.D = \frac{\sum d}{N}$$

$\sum d$ = Calculated regardless the sign in the previous example.

A.D 1 = 6/4 = 1.5 A.D 2 = 20/4 = 5

3] Variance (S^2)

❑ It is the mean squared deviations of the values from their mean divided by the number of observation minus 1.

$$S^2 = \frac{\sum(X - X^-)^2}{N-1} = \frac{\sum d^2}{N-1}$$

- ❑ S^2 = Variance
- ❑ X^- = mean
- ❑ X = individual value
- ❑ N = number of observations
- ❑ d^2 = Squared deviation of the data from their arithmetic mean.

Value X	Mean X^-	d	d^2	$\sum d^2$
20	23	3	9	20
22	23	1	1	
24	23	1	1	
26	23	3	9	

S^2 = 20/3 = 6.67

4] Standard deviation S or S.D

- ❑ It is the most commonly used measure of variability.
- ❑ It is the square root of the variance.
- ❑ N – 1 = called the degree of freedom (D.F.).
  - ↳ If the observation is sufficiently large (more than 30) we can use N instead of N- 1 because the product dose not differ significantly in this case.
  - ↳ In case of much variation the S.D. is high i.e. the higher S.D. the more spread out measured.

$$S.D. = \sqrt{Variance} \qquad S.D. = \sqrt{\frac{\sum d^2}{N-1}}$$

Mean is valid if: it is ≥ 2.5 x S.D

Calculate the standard deviation of the following data: 20-22-24-26

Value X	Mean X^-	d	d^2	$\sum d^2$
20	23	3	9	20
22	23	1	1	
24	23	1	1	
26	23	3	9	

$$S.D = \sqrt{\frac{20}{4-1}} = 2.58$$

2.5 x 2.58= 6.45

X^- = 23 > 6.45 ⇨ i.e. mean is valid

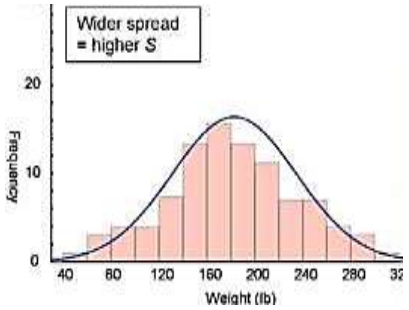
Standard deviation and variance

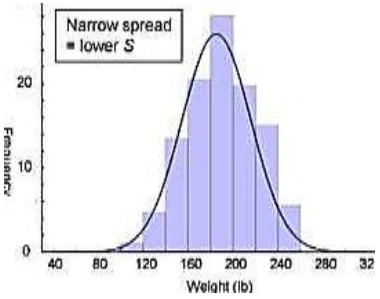
The standard deviation describes the "average distance" of data value from their mean.

Xi (individual measurement)	Xi – X = di	$(di)^2$
10	0.8	0.64
12	1.2	1.44
8	-1.2	1.44
14	3.2	10.44
10	0.8	0.64
Total = 54		14

Mean = 54/5 = 10.8 Variance = $S^2 = (di)^2/N-1 = 14/4 = 3.6$ S.D = $\sqrt{3.6}$ = 1.9

Wide spread results in higher SDs narrow spread in low SDs





5] Coefficient of variation (C.V.)

- Used to facilitate comparison between data sets with different units (if we have a set of data whose unit in Kg, this can't be compared directly with another set of data in cm).

$$C.V = \frac{S.D}{\bar{X}} \; x \; 100$$

Example 1	Example 2
❑ 2 drugs were bio assayed for their action on the body weight of 2 groups of rats. ❑ The mean body weight of the first group was 50 gm with a S.D. 4.5, and the second group was 75 gm with S.D. 5.5. ❑ Compare the degree of dispersion of those figures around their means in the 2 groups.	❑ In a screening of a new drug by the isolated rabbit heart, the mean value of the heart rate was 70 beat/min and the S.D. was 9.5. ❑ The % of catecholamine uptake was 46.21 with S.D. of 4.24. ❑ Compare the variability of the result of the H.R. and catecholamine uptake
C.V. of 1st group = 4.5/50 x 100 = 9 C.V. of 2nd group = 5.5/75 x 100 = 7.3 ♣ This means that the body weight of 1st group is more dispersed around their mean than the 2nd group.	C.V. H.R. = 9.5/70 X 100 = 13.7 C.V. C U = 4.2/46.2 X 100 = 9.17 ♣ This means that the heart rate is more dispersed around their mean that the catecholamine uptake

6] Standard Error (S.E)

- ❑ It describes the variability of the mean.
- ❑ It is the standard deviation of the mean.

$$S.E = \frac{S.D}{\sqrt{N}} = \sqrt{\frac{\sum d^2}{n\,(n-1)}}$$

Mean is valid if: it is ≥ 10 x S.E

♣ Calculate the S.E. of the following data: 20-22-24-26, S.D = 2.58, $\bar{X}$ = 23.

$$S.D = \frac{2.58}{\sqrt{4}} = 1.29$$

Value	Mean	Deviation d	Squared d^2	Σd^2
26	23	3	9	20
22	23	1	1	
24	23	1	1	
20	23	3	9	

$10 \times S.E. = 10 \times 1.29 = 12.29$

the mean is valid "23 > 12. 9"

الدكتوره اتكلمت عن ال**Test of significance** في اخر دقيقه وقالت ان الدكتورحمدى هيكمله معاكم

Tests of significance

Purpose

- Tests of significance are used to determine whether the difference between two or more values is significant or not.

Significant difference	Means that it will still persist when experiment is repeated several times and will not be zero at any time.
Non-significant difference	Means that the difference can occur normally and that difference can occur by chance and not due to an external factor such as the effect of a drug or due to a disease.

Notes

- The tests of significance depend on the fact that there are factors or items which are calculated to determine the limits of the normal variations that occur by chance.
- These factors are given in tables and their values depend on the number of experiments in the series of observations. Ex: Z, t, X^2 and F values.
- These values are calculated from the results of the experiment and the calculated value is compared with the tabulated value.
  - ✓ If calculated value > tabulated value ➔ significant difference
  - ✓ If calculated value < tabulated value ➔ No significant difference

End of Mid-term part (4 lectures)